Markov Chains, Stochastic Processes, and Advanced Matrix Decomposition

Jack Gilbert

Introduction

This paper seeks to explore, in some detail, the basics of stochastic processes and their extensive relationship with Markov chains. This is primarily a study of the field of statistics. However, by broadening our scope to consider techniques outside the realm of statistics, we can utilize advanced techniques from linear algebra (and some basic ones, too!), thus broadening and enhancing our understanding of these topics.

Probability Spaces

We proceed with our study of Markov chains by defining some of the basics of probability spaces. Understanding these basic definitions is crucial to understanding the more complicated ideas presented later in the paper, especially in our observations of matrix decompositions. Lets begin.

Definition 1. *An event is a subset of a sample space. An event A is said to occur if and only if the observed outcome $\omega \in A$.*

Definition 2. *Given a sample space Ω and an event A, the events complement A^c is the event which occurs if and only if A does not occur. $A^c = \{\omega \in \Omega \| \omega \notin A\}$.*

Definition 3. *If Ω is a sample space and if P is a function which associates a number for each event in Ω, then P is called the probability measure provided that*

a. for any event $A, 0 \leq P(A) \leq 1$
b. $P(\Omega) = 1$
c. For any sequence $A_1, A_2, \ldots$ of disjoint events, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Random Variables and Expected Outcomes

We will now shift our focus to the study of random variables, a key component of the study of stochastic processes.

Definition 4. *A random variable X with values in the set E is a function which assigns a value $X(\omega) \in E$ to each outcome $\omega \in \Omega$.*

When E is finite, X is said to be a discrete random variable.

Definition 5. *The discrete random variables $X_1, \ldots, X_n$ are said to be independent if $P\{X_1 = a_1, \ldots, X_n = a_n\} = P\{X_1 = a_1\} \cdots P\{X_n = a_n\}$ for every $a_1, \ldots, a_n \in E$.*

Expected values are the potential outcomes for each random variable. The expected value for a random variable X taking values in the set $E \subset \mathbb{R}_+$ is

$$E[X] = \sum_{a \in E} aP\{X = a\}.$$

With these definitions in place, we can proceed to define Markov chains, which are the primary focus of this paper.

Markov Chains

We now begin our study of Markov chains.

Definition 6. *The stochastic process $X = \{X_n; n \in \mathbb{N}\}$ is called a Markov chain if $P\{X_{n+1} = j \| X_0, \ldots, X_n\} = P\{X_{n+1} = j \| X_n\}$ for every $j \in E, n \in \mathbb{N}$.*

So a Markov chain is a sequence of random variables such that for any n, X_{n+1} is conditionally independent of $X_0, \ldots, X_{n-1}$ given X_n. We use

$$P\{X_{n+1} = j \| X_n = i\} = P(i, j) \text{ where } i, j \in E$$

is independent of n. The probabilities $P(i, j)$ are called the transition probabilities for the Markov chain X. The Markov Chain is said to be time homogenous.

We can arrange the $P(i, j)$ into a square matrix P called the transition matrix of the Markov chain X. For example, for the set $E = \{1, 2, \ldots\}$ the transition matrix is

$$P = \begin{bmatrix} P(0,0) & P(0,1) & P(0,2) & \cdots \\ P(1,0) & P(1,1) & P(1,2) & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix} \tag{1}$$

The basic notation we will use with regard to these matrices is as follows:
$M(i, j)$ refers to the entry in row i, column j of matrix M.

Column vectors are represented by letters, i.e. $f(i)$ is column fs i-th entry.

Row vectors are greek letters, i.e. $\pi(j)$ is row πs j-th entry.

Definition 7. *Let P be a square matrix with entries $P(i, j)$ where $i, j \in E$. P is called a Markov matrix over E if*

a. For every $i, j \in E, P(i, j) \geq 0$

b. For every $i \in E, \sum_{j \in E} P(i, j) = 1$.

We can now present our first theorem.

Theorem 1. *For every $n, m \in \mathbb{N}$ with $m \geq 1$ and $i_0, \ldots, i_m \in E$,*

$$P\{X_{n+i} = i, \ldots, X_{n+m} = i_m \| X_n = i_0\} = P(i_0, i_1)P(i_1, i_2) \cdots P(i_{m-1}, i_m).$$

The following corollary provides more insight into this theorem.

Corollary 1. *Let π be a probability distribution on E. Suppose $P\{X_0 = i\} = \pi(i)$ for every $i \in E$. Then for every $m \in \mathbb{N}$ and $i_0, \ldots, i_m \in E$,*

$$P\{X_0 = i_0, X_1 = i_1, \ldots, X_m = i_m\} = \pi(i_0)P(i_0, i_1) \cdots P(i_{m-1}, i_m).$$

Theorem 1 and Corollary 1 show that the joint distribution of $X_0, \ldots, X_m$ is completely specified for every m once the initial distribution π and transition matrix P are known. We can get the joint distribution of $X_{n1}, \ldots, X_{nk}$ for any integer $k \geq 1$ and $n_1, \ldots, n_k \in \mathbb{N}$. This leads us to the following proposition.

For any $m \in \mathbb{N}$,

$$P\{X_{n+m} = j \| X_n = i\} = P^m(i, j) \text{ for every } i, j \in E \text{ and } n \in \mathbb{N}.$$

In other words, the probability that the chain moves from state i to that j in m steps is the (i, j) entry of the n-th power of the transition matrix P. Thus for any $m, n \in \mathbb{N}$,

$$P^{m+n} = P^m P^n$$

which in turn becomes

$$P^{m+n}(i, j) = \sum_{k \in E} P^m(i, k)P^n(k, j); i, j \in E.$$

This is called the Chapman-Kolmogorov equation. This states that when starting at state i, in order for process X to be in state j after $m + n$ steps, it must be in some intermediate state k after the m-th step and then move from state k into state j in the remaining n steps. This is illustrated in the following example.

Example 1. *Let $X = \{X_n; n \in \mathbb{N}\}$ be a Markov chain with state space $E = \{a, b, c\}$ and transition matrix*

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{bmatrix} \tag{2}$$

Then

$$\begin{aligned} & P\{X_1 = b, X_2 = c, X_3 = a, X_4 = c, X_5 = a, X_6 = c, X_7 = b \| X_0 = c\} && (3) \\ & = P(c, b)P(b, c)P(c, a)P(a, c)P(c, a)P(a, c)P(c, b) && (4) \\ & = \frac{2}{5} * \frac{1}{3} * \frac{3}{5} * \frac{1}{4} * \frac{3}{5} * \frac{1}{4} * \frac{2}{5} && (5) \\ & = \frac{3}{2500}. && (6) \end{aligned}$$

The two-step transition probabilities are given by

$$P = \begin{bmatrix} \frac{17}{30} & \frac{9}{40} & \frac{5}{24} \\ \frac{8}{15} & \frac{3}{10} & \frac{1}{6} \\ \frac{17}{30} & \frac{3}{20} & \frac{17}{60} \end{bmatrix} \tag{7}$$

where in this case $P\{X_{n+2} = c \| X_n = b\} = P^2(b, c) = \frac{1}{6}$.

State Spaces

With an understanding of the Chapman-Kolmogorov equation as the basis of our study of Markov chains and Markov matrices we can move on to our classification of the various states we will encounter throughout this paper. We will define these states now.

Definition 8. *Given a Markov chain* X*, a state space* E*, a transition matrix* P*, let* T *be the time of the first visit to state* j *and let* N_j *be the total visits to state* j*. Then*

a. j *is recurrent if* $P_j\{T < \infty\} = 1$*. Otherwise,* j *is transient if* $P_j\{T = +\infty\} > 0$

b. A recurrent state j *is null if* $E_j[T = \infty$; otherwise j is non-null

c. A recurrent state j is periodic with period δ if $\delta \geq 2$ is the largest integer for $P_j\{T = n\delta$ for some $n \geq 1\} = 1$.

So if j is recurrent, $F(j, j) = 1$ so the probability of returning to state j is 1. If j is transient, $F(j, j) < 1$ and there is a positive probability of never returning to state j. This leads us to our next theorem.

Theorem 2. *If* j *is transient or recurrent null, then for every* $i \in E, \lim_{n\to\infty} P^n(i, j) = 0$*. If* j *is recurrent non-null aperiodic, then* $\pi(j) = \lim_{n\to\infty} P^n(j, j) > 0$ *and for every* $i \in E, \lim_{n\to\infty} P^n(i, j) = F(i, j)\pi(j)$*.*

So if j is periodic with period δ then returning to j is only possible on steps $\delta, 2\delta, 3\delta, \ldots$ and the same is true of each successive return to j. Hence

$$P^n(j, j) = P_j\{X_n = j\} > 0 \text{only if} n \in \{0, \delta, 2\delta, \ldots\}.$$

In order to tell whether $P^n(j, j) > 0$ or not, we do not need to calculate $P^n(j, j)$. Instead we say j can be reached from i, $i \to j$, if there exists an integer $n \geq 0$ such that $P^n(i, j) > 0$. Thus for $i \neq j, i \to j$ if and only if $F(i, j) > 0$. In order for $i \to j$ there must be a sequence $i_1, i_2, \ldots, i_n$ of states such that $P(i, i_1) > 0, P(i_1, i_2) > 0, \ldots, P(i_n, j) > 0$. Otherwise $P^n(i, j) = 0$ for every n and thus it is impossible for $i \to j$.

Definition 9. *a. A set of states is closed if no state outside the set can be reached from within the set*

b. A state forming a closed set by itself is called an absorbing state

c. A closed set is irreducible if no proper subset of it is closed

d. A Markov chain is irreducible if its only closed set is the set of all states.

Example 2. *Consider the Markov chain with state space* $E = \{a, b, c, d, e\}$ *and transition matrix*

$$P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix} \tag{8}$$

The closed sets are $\{a, b, c, d, e\}$ *and* $\{a, c, e\}$*. Since there exist two closed sets, the chain is not irreducible. By deleting the second and fourth rows and column we end up with the matrix*

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \tag{9}$$

which is the Markov matrix corresponding to the restriction of X *to the closed set* $\{a, c, e\}$*. We can rearrange* P *for easier analysis as such:*

$$P^* = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Recurrence and Irreducible Markov Chains

We have a lemma which shows that if j is recurrent and $j \to k$, then $k \to j$ and $F(k, j) = 1$. This should be intuitive based on our definition of a recurrent state: we know since j is recurrent, $F(j, j) = 1$, so if $j \to k$ then it follows that $k \to i$ must at some point return to $j, i \to j$. This notion of recurrence gives us the basis of the following theorem.

Theorem 3. *From a recurrent state, only recurrent states can be reached.*

From this theorem we know that no transient state can be reached from any recurrent state and thus the set of all recurrent states is closed. This leads us to one of the main results of Markov matrices. First, a lemma, and then our main theorem:

Lemma 1. *For each recurrent state* j *there exists an irreducible closed set* C *which includes* j*.*

Proof. Let j be a recurrent state and let C be the set of all states which can be reached from j. Then C is a closed set. We show if $i, k \in C$ then $i \to k$:

If $i \in C$ then $j \to i$. Since j is recurrent, our previous lemma implies that $i \to j$. There must be some state k such that $j \to k$, and thus $i \to k$. □

We will use this lemma in formulating the following theorem.

Theorem 4. *In a Markov chain, the recurrent states can be divided uniquely into irreducible closed sets $C_1, C_2, \ldots$*

The proof follows directly from the lemma preceding it. In addition to $C = C_1 \cup C_2 \cup \ldots$ of recurrent states, the chain includes transient states as well (since recurrent states can be reached via transient states). Using this theorem we can arrange our transition matrix in the following form:

$$P = \begin{bmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ 0 & 0 & P_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & vdots & \vdots \\ Q_1 & Q_2 & Q_3 & \cdots & Q \end{bmatrix} \tag{10}$$

where $P_1, P_2, \ldots$ are the Markov matrices corresponding to sets $C_1, C_2, \ldots$ of states. Each one of these sets is an irreducible Markov chain. The following theorem ties this all together.

Theorem 5. *Let X be an irreducible Markov chain. Then either all states are transient, or all are recurrent null, or all are recurrent non-null. Either all states are aperiodic, or else all are periodic with the same period δ.*

The following corollaries conclude this section of the paper.

Corollary 2. *Let C be an irreducible closed set with finitely many states. Then no state in C is recurrent null.*

Corollary 3. *If C is an irreducible closed set with finitely many states, then C has no transient states.*

In the case with finitely many states we now have all the tools necessary to classify the states! We first identify the irreducible closed sets, then use our theorem and corollaries to conclude that all states belonging to an irreducible closed set are recurrent non-null; the remaining states, if any, are transient; and periodicity is determined by again applying our theorem.

Markov Chains and Linear Algebra

How does this study of Markov chains and transition matrices relate to linear algebra? First, a theorem to give you some of the flavor:

Theorem 6. *Let X be an irreducible Markov chain. Consider the system of linear equations*

$$\vec{v}(j) = \sum_{i \in E} \vec{v}(i)P(i,j), j \in E.$$

Then all states are recurrent non-null if and only if there exists a solution $\vec{v}$ with $\sum_{j \in E} \vec{v}(j) = 1$.

If there is a solution $\vec{v}$ then $\vec{v}(j) > 0$ for every $j \in E$, and $\vec{v}$ is unique.

We now fully engage with linear applications of Markov chains and stochastic matrices.

The Perron-Frobenius Theorem

We can use the general case of a theorem proved by a mathematician named Perron in 1907 and extended by Frobenius in 1912 to obtain a corollary with some interesting properties pertaining to our understanding of Markov chains. The proof of this theorem has been omitted, as it favors an analysis approach which would not be suited for this paper; however, its corollary will be crucial to our discussion.

Theorem 7. *Perron-Frobenius part 1:*

Let A be a square matrix of size n with non-negative entries. Then

1. A has a positive eigenvalue λ_0 with left eigenvector $\vec{x_0}$ such that $\vec{x_0}$ is non-negative and non-zero
2. If λ is any other eigenvalue of $A, \|\lambda\| \leq \lambda_0$
3. If λ is an eigenvalue of A and $\|\lambda\| = \lambda_0$, then $\mu = \frac{\lambda}{\lambda_0}$ is a root of unity and $\mu^k \lambda_0$ is an eigenvalue of A for $k = 0, 1, 2, \ldots$

Theorem 8. *Perron-Frobenius part 2:*

Let A be a square matrix of size n with non-negative entries such that A^m has all positive entries for some m. Then

1. A has a positive eigenvalue λ_0 with a corresponding left eigenvector $\vec{x_0}$ where the entries of $\vec{x_0}$ are positive
2. If λ is any other eigenvalue of $A, \|\lambda\| < \lambda_0$
3. λ_0 has multiplicity 1.

We will now give the corollary related to the Perron-Frobenius theorems.

Corollary 4. *Let P be an irreducible Markov matrix. Then 1 is a simple eigenvalue of P. For any other eigenvalue λ of P we have $\|\lambda\| \leq 1$. If P is aperiodic then $\|\lambda\| < 1$ for all other eigenvalues of P. If P is periodic with period δ then there are δ eigenvalues with an absolute value equal to 1. These are all distinct and are*

$$\lambda_1 = 1, \lambda_2 = c, \ldots, \lambda_\delta = c^{\delta - 1}; c = e^{2\pi i/\delta}.$$

Here well use an example which will reveal some of the more interesting results of these theorems and their corollary.

Example 3.

$$P = \begin{bmatrix} 0 & 0 & 0.2 & 0.3 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \end{bmatrix} \tag{11}$$

P is irreducible and periodic with period $\delta = 2$. Then

$$P^2 = \begin{bmatrix} 0.48 & 0.52 & 0 & 0 & 0 \\ 0.70 & 0.30 & 0 & 0 & 0 \\ 0 & 0 & 0.38 & 0.42 & 0.20 \\ 0 & 0 & 0.20 & 0.30 & 0.50 \\ 0 & 0 & 0.44 & 0.46 & 0.10 \end{bmatrix} \tag{12}$$

The eigenvalues of the 2 by 2 matrix in the top-left corner of P^2 are 1 and -0.22. This is sufficient enough to compute all eigenvalues of P by using the Perron-Frobenius theorems, which imply that the set of eigenvalues of a periodic irreducible matrix P, when regarded as a system of points in the complex plane, goes over into itself under a rotation of the plane by the angle $2\pi/\delta$. Wow! Since $1, -0.22$ are eigenvalues of P, their square roots will be eigenvalues for P : $1, -1, i\sqrt{0.22}, -i\sqrt{0.22}$. The final eigenvalue must go into itself by a rotation of 180 degrees and thus must be 0.

It is worth taking the time to pause and consider an idea that has been implicit in several of these theorems. We will now take the time to make this explicit:

Theorem 9. *All finite stochastic matrices P have 1 as an eigenvalue and there exist non-negative eigenvectors corresponding to $\lambda = 1$.*

We will use this theorem throughout the remainder of the paper, so it is worth proving.

Proof. Since each row of P sums to 1, $\vec{y}$ is a right eigenvector. We try to find a left eigenvector with non-negative components that sum to 1:

Since all finite chains have at least one positive persistent state, we know there exists a closed irreducible subset S and the Markov chain associated with S is irreducible positive persistent. We know for S there exists an invariant probability vector. Assume P is a square matrix of size n and rewrite P in block form with

$$P* = \begin{bmatrix} P_1 & 0 \\ R & Q \end{bmatrix} \tag{13}$$

where P_1 is the probability transition matrix corresponding to S. Let $\vec{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ be the invariant probability vector for P_1. Define $\vec{\gamma} = (\pi_1, \pi_2, \ldots, \pi_k, 0, 0, \ldots, 0)$ and note $\vec{\gamma}P = \vec{\gamma}$. Hence $\vec{\gamma}$ is a left eigenvector for P corresponding to $\lambda = 1$. Additionally,

$$\sum_{i=1}^{n} \gamma_i = 1.$$

□

This theorem shows that $\lambda = 1$ is the largest possible eigenvalue for a finite stochastic matrix P. This fact is related to the irreducibility of the Markov chain determined by P, which we will show in the next theorem.

Theorem 10. *If P is the transition matrix for a finite Markov chain, then the multiplicity of the eigenvalue 1 is equal to the number of irreducible subsets of the chain.*

Proof. The first half of this proof goes as follows. Arrange P in block form based on the irreducible subsets of the chain $C_1, C_2, \dots$ as before.

$$P = \begin{bmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & P_m & \vdots \\ R_1 & R_2 & \cdots & R_m & Q \end{bmatrix} \tag{14}$$

Each P_k corresponds to a subset C_k for an irreducible positive persistent chain. The $\vec{x_i}$s for each C_i are linearly independent; thus the multiplicity for $\lambda = 1$ is at least equal to the number of subsets C. □

Advanced Matrix Decomposition and Markov Chains

The last topic we will examine in this paper deals with matrix decompositions and Markov chains. We will start by looking at infinite stochastic matrices. We can apply techniques of linear decomposition to infinite stochastic matrices to characterize matrices with some interesting stochastic properties. We will begin with a theorem.

Theorem 11. *Let $\{X_t; t = 0, 1, 2, \dots\}$ be a Markov chain with state $\{0, 1, 2, \dots\}$ and a transition matrix P. Let P be irreducible and consist of persistent positive recurrent states. Then*

$$I - P = (A - I)(B - S)$$

where A is strictly upper triangular with $a_{ij} = E_i[\text{number of times } X = j \text{ before } X \text{ reaches } \Delta_{j-i}], i < j$;

B is strictly lower triangular with $b_{ij} = P_i\{X_1^i = j\}, i > j$;

and S is diagonal where $s_j = \sum_{j=0}^{i-1} b_{ij} (\text{and } s_0 = 0)$.

Moreover, $a_{ij} < \infty$ and $i < j$.

This primary result can then be used to find the stationary distribution, the matrix of mean first-passage times, and the fundamental matrix for discrete-time Markov chains. How nice!

The final result we will examine takes the previous theorem and extends it to for null-recurrent and transient cases.

Theorem 12. *Define E_D to be the diagonal matrix with entries e_i on the diagonal where $e_i = P_i(X_n \notin \{0, 1, 2, \dots\})$ for $n \geq 1$. Then for every infinite irreducible stochastic matrix P with A, B and entries a_{ij}, b_{ij} respectively,*

1. *$I - P = (A - I)(B - S) + E_D$*

2. *P is recurrent if and only if $I - P = (A - I)(B - S)$, or in other words, if $E_D = 0$.*

Spectral Representations

Suppose we have a diagonalizable matrix A. Define B_k to be the matrix obtained by multiplying the column vector f_k with the row vector π_k where f_k, π_k are from A. Explicitly,

$$B_k = f_k \pi_k.$$

Then we can represent A in the following manner:

$$A = \lambda_1 B_1 + \lambda_2 B_2 + \cdots + \lambda_n B_n.$$

This is the spectral representation of A, and it holds some key properties relevant to our discussion of Markov chains. For example, for the k-th power of A we have

$$A^k = \lambda_1^k B_1 + \cdots + \lambda_n^k B_n.$$

This provides the means for computing the k-step transition matrix of A. This may then be used to obtain the limits of A^k as $k \to \infty$, along with estimates of the rate of convergence and bounds for error involved in certain calculations. An example will highlight some of these properties:

Example 4. *Let*

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} \tag{15}$$

We get from $P1 = 1$ that $\lambda_1 = 1$ and $f_1 = 1$ is its eigenvector. Since the trace of P is 1.5 and the sum of the eigenvalues of P equal the trace of P, λ_2 must be 0.5. We can now calculate B_1:

$$B_1 = f_1 \pi_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} \tag{16}$$

and since $P^0 = I = B_1 + B_2$ for $k = 0$,

$$B_2 = \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix} \tag{17}$$

Thus the spectral representation for P^k is

$$P^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} + (0.5)^k \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix}, k = 0, 1, \ldots \tag{18}$$

The limit as $k \to \infty$ has $(0.5)^k$ approach zero, so

$$P^\infty = \lim_k P^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} \tag{19}$$

Using spectral representations and given certain conditions we can relatively easily calculate powers of P. This is done via a process quite similar to diagonalization. We begin by first applying a theorem which holds for general finite matrices, and then we apply our result to stochastic matrices.

Theorem 13. *Let A be a square matrix of size n with n linearly independent eigenvectors $\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}$ and associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Define the matrices L, D by*

$$L = \begin{bmatrix} \vec{x_1} \\ \vec{x_2} \\ \vdots \\ \vec{x_n} \end{bmatrix}$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{20}$$

Then $A = L^{-1}DL$.

This can be applied for size n stochastic matrices. If P contains a set n of linearly independent left eigenvectors then P^k can be determined from $L^{-1}DL$.

Conclusion

Markov chains and their affiliated transition matrices have a wide variety of application both within the field of math and in real-world application. By understanding how linear algebra interacts with Markov chains, one gains a deeper intuition as to how stochastic matrices can be manipulated while still retaining their statistical properties in such a way that proves useful.

Bibliography

[1] Robert Beezer. A Second Course In Linear Algebra. Open-License, 2014.

[2] Erhan Cinlar Introduction to Stochastic Processes. Prentice-Hall, Inc., first edition, 1975.

[3] Daniel Heyman A Decomposition Theorem for Infinite Stochastic Matrices. J. Appl. Prob. 32, 893-901, 1995.

[4] Yiqiang Zhao, Wei Li, W. John Braun On a Decomposition for Infinite Transition Matrices. 1997.

[5] Dean L. Isaacson, Richard W. Madsen Markov Chains: Theory and Applications. John Wiley Sons, first edition, 1976.